

## The value of indirect ties in citation networks: SNA analysis with OWA operator weights

Article (Accepted Version)

Marra, Marianna, Emrouznejad, Ali, Ho, William and Edwards, John S (2015) The value of indirect ties in citation networks: SNA analysis with OWA operator weights. Information Sciences, 314. pp. 135-151. ISSN 0020-0255

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/89741/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

# The value of indirect ties in citation networks: SNA analysis with OWA operator weights

<sup>a</sup> Marianna Marra, <sup>a</sup> Ali Emrouznejad, <sup>b</sup> William Ho, <sup>a</sup> John S. Edwards

<sup>a</sup> *Operations and Information Management Group, Aston Business School, Aston University, Birmingham, UK*

<sup>b</sup> *Department of Management and Marketing, The University of Melbourne, Level 10, 198 Berkeley Street, Victoria 3010, Australia*

## Abstract

This paper seeks to advance the theory and practice of the dynamics of complex networks in relation to direct and indirect citations. It applies social network analysis (SNA) and the ordered weighted averaging operator (OWA) to study a patent citations network. So far the SNA studies investigating long chains of patents citations have rarely been undertaken and the importance of a node in a network has been associated mostly with its number of direct ties. In this research OWA is used to analyse complex networks, assess the role of indirect ties, and provide guidance to reduce complexity for decision makers and analysts. An empirical example of a set of European patents published in 2000 in the renewable energy industry is provided to show the usefulness of the proposed approach for the preference ranking of patent citations.

**Keywords:** Social network analysis (SNA); patent citations network; aggregation operator; ordered weighted averaging (OWA);

## Introduction

Work on citation networks has been increasing (Newman, 2001; Batagelj, 2003; Karrer and Newman, 2009; Radicchi et al., 2012), and the growing interest in different network measures is based on their impact on our understanding of the knowledge diffusion process in disciplines (in the case of academic citation networks) and technological innovation (in the case of patent citation networks). Citation networks have several important features, including showing the relation between number of citations and time. For example, the number of citations received by a node (paper or patent) decreases with age, and the number of citations to a given node is considered a good estimate of its relevance and prestige within the network. Citation networks are directed and essentially acyclical.

We know that knowledge flows from one node to another; thus, the influence of previous nodes on a citation path can be considered important for understanding the

importance of citation network nodes. Studies employing social network analysis (SNA) to analyse citation networks usually measure network centrality by considering direct ties (Whitley and Galliers, 2007; Chang et al., 2009; Lee et al., 2010). Network centrality measures the number of each node's connections, and the number of ties is an indication of the importance of the network node (Borgatti and Everett, 2006). Some studies employ specific algorithms to map citation networks and understand the flows of knowledge across them. However, very few studies investigate more than three generations of citations (von Wartburg et al., 2005; Trajtenberg et al. 2002); examining patent citations over several generations could enrich our understanding of citation network dynamics.

This paper contributes by ranking patent citations using Ordered Weighted Averaging (OWA) (Emrouznejad and Amin, 2010), with the aim of obtaining a score that explains the longevity of patents over time. This approach provides a better explanation of patent success than SNA alone. The basic idea is that the diffusion process in directed networks is explained better by considering the indirect citations received over time than by relying on purely local measures such as citation counts. Analysis of indirect ties sheds light on otherwise underestimated aspects of citation networks. We show how information and knowledge flow between a network's nodes.

The proposed OWA operator weights proposed by Emrouznejad and Amin (2010), can be used for preference ranking aggregation. In the present study we employ their formulation, assuming a number of patents and the corresponding number of direct and indirect citations, to estimate a score for each patent. These scores should reflect the impact of direct and indirect citations on patent life.

The remainder of the paper is organized as follows. Section 2 discusses the literature on patent citations data. Section 3 focuses on one SNA algorithm, the Hub and Authorities, used for citation network analysis. Section 4 discusses the OWA method and Section 5 describes the data. Section 6 presents the results of application of these SNA algorithms to patent data. Section 7 discusses the results and offers some conclusions. There is also a supplement document that shows details of the results on each node, this document is available on request.

## **Patent citation networks: Structural characteristics**

### Patent citations

The increase in international patenting activity has resulted in increased use of patent data in research on technological change and innovation, to capture aspects of successful product innovation in firms, and the spread of technologies over time. Patent data are popular because of (a) their availability and (b) their utility as technology indicators (Jaffe et al., 1993; Chang et al., 2009). The combination of patent citations and SNA has been employed in several studies to assess the importance, radicalness and novelty of patents and inventions. Patent citations are treated as the links between patents and allow an understanding of knowledge flows across countries and technological fields. The network of patent citations shows the direction of the innovation process and, thus, provides a good representation of the way that incremental and radical innovation occur over time. An increasing stream of research is exploiting patent citations and SNA techniques to analyse and investigate citation networks in order to make inferences about the value of patents (Trajtenberg, 1990; Bacchiocchia and Montobbio, 2004; Duguet and MacGarvie, 2005; Batagelj et al., 2006; Cantner and Graf 2006).

Mina et al. (2007) apply these tools to develop a longitudinal analysis of a large citation network related to a bibliographic database of 11,240 papers and 5,136 US patents in the medical sector, to show the mechanisms through which medical knowledge emerges, grows and is transformed. Whitley and Galliers (2007) adopt a network-based approach to studying the citation networks of papers published in the field of information systems, to identify core themes. Fontana et al. (2009) use patent citation networks to study the evolution of data communications technology. Barbera et al. (2011) apply SNA algorithms to the patent citation network of a single product, the artificial intervertebral disc, to show its scientific development in the US. Choi et al. (2011) analyse keyword networks in management information system research to understand the evolution of knowledge in the field. However, although several studies use patent citation counts to assess the importance of inventions, few focus on the longevity and survival of these innovations over time on the basis of indirect links/citations. For example, approaches that emphasize the spread of technology tends to underestimate these aspects.

A patent is a detailed document that includes a set of exclusionary rights granted by a state to an inventor or assignee. The information in a patent document includes,

amongst other things, references to previous patents and the scientific literature. Patent references fulfil a specific legal function and follow different rules to journal article references. Journal article citations are introduced only by the article's author(s); patent citations are added by both patent applicant (inventor) and the patent examiner.

All patents contain references to prior patents, but citation practices differ across patent offices. Patent applications to the United States Patent and Trademark Office (USPTO) must include citations to previous patents related to the new invention, but this is not compulsory for European Patent Office (EPO) patent applications. Nevertheless, inventors applying to the EPO will generally include all relevant details in the search report in order to avoid future claims and objections.

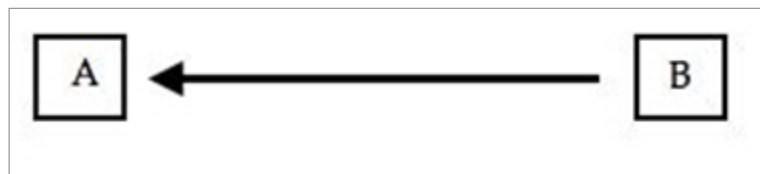
Most studies examine the relationship between patents and their importance, making two main assumptions. The first is that citations received by a patent (forward citations) are a good measure of the cited patent's importance: the more important, the more often it will be cited. In other words, forward citations are an indicator of the diffusion of a technology and its application and value over time, and also are a good measure of the importance of the focal patent for the future development of the technological domain, and the patent's potential economic value (Harhoff et al., 2003; Hall et al., 2005). It is generally accepted that more important patents will tend to be cited more than less important ones. Thus, the second assumption is that the citations included in a patent (backward citations) are a measure of the novelty of the innovation and knowledge domain on which it relies. Citations to previous patents are an indication that the citing patent, to some extent, relies on those previous innovations, and that they have been useful for developing the new knowledge (Jaffe et al., 1993). Backward citations can refer to other patents or to non-patent literature, such as scientific papers, and generally are considered to be positively related to the value of the focal patent (Harhoff et al., 2003) and its novelty or radicalness (Carpenter et al., 1981; Ahuja and Morris Lampert, 2001; Rosenkopf and Nerkar, 2001; Dahlin and Behrens, 2005).

In both cases, the researcher focuses on direct citations between patents, which under-estimates the role of the indirect ties between them. Both forward and

backward citations can be used to construct the citation network for a sample of patents, using SNA.

#### Citation network

A patent citation network refers to a directed graph  $N = (V, L)$  where  $V$  is a set of vertices and  $L$  is a set of arcs. The following condition describes a citation network: the arc  $(v, u)$  goes from vertex  $v \in V$  to vertex  $u \in V$  if the patent  $(v)$  cites the patent  $(u)$ . The arrow in Figure 1 represents the flow of knowledge from a node or patent (A) to another node or patent (B). The citation made by B to A is a backward citation. From the receiver's (B) point of view, it is a forward citation.



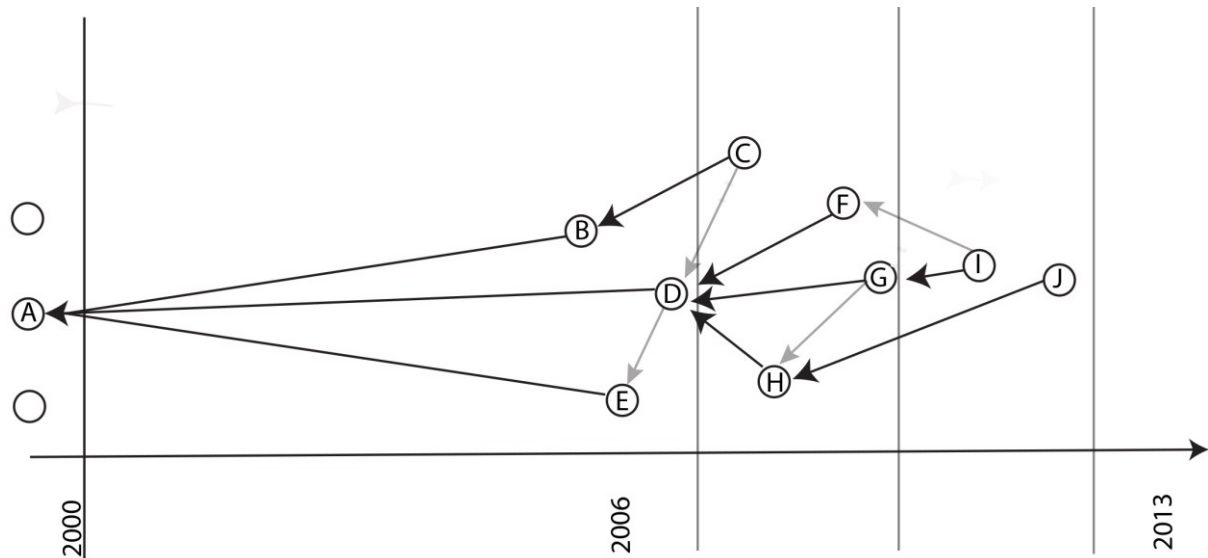
**Figure 1. Example of a patent and citation**

Citation networks have four main properties:

- acyclicity which refers to the time dimension - the patent can cite only previous patents;
- directionality which refers to the direction of ties – arcs indicate the direction of knowledge flows;
- irreflexivity which means no patent can cite itself;
- direct and indirect links.

These properties are displayed in Figure 2, which shows the forward citations received by A. Let A be a patent published in 2000 and B, D and F be patents published around 2006, and citing A. The links between A and B, A and D, and A and F are direct citations. If C cites B without citing A this is an indirect citation. C also cites D, indicated by the grey line, which means that although it might appear to be another indirect citation it is not considered as such because of the previous direct relation between D and A. In other words, as there are two or more indirect citations, but all refer to the same original node, we count them only once. While it is straightforward to identify direct citations, indirect citations require that each node in the network is counted only once and, more specifically, on its first appearance. Studies on citation networks consider indirect citations in terms of co-citation and

bibliographic coupling (Garfield 1979, Small, 1973; Kessler, 1963). Co-citation is defined as the edge between two nodes cited by the same node(s). In Figure 2, nodes B and D are cited by C. As Boyack and Klavans (2010) noted, co-citation analysis is used in mapping science to identify the research front within a discipline. Bibliographic coupling is defined as the edge between two nodes citing the same node(s). In Figure 2, there is bibliographic coupling between F and G, as they both cite D. Recent studies have validated the performance of these methods to detect emerging research front (Shibata et al. 2009; Boyack and Klavans, 2010) and investigated citation networks with combinational types of citations, such as including direct citations and co-citations to exploit new possibilities of detecting research fronts (Fujita et al. 2014). The concept of indirect citations proposed in this paper differs from these studies, as described in our example referring to Figure 2.



**Figure 2. Example of a citation network with direct and indirect citation**

In Figure 3, the same concept is re-drawn to show better that the nodes along the diagram that appear more than once are counted only at their first appearance in time; subsequent appearances of the same node are coloured grey to indicate that they have not been counted. 1st place refers to direct citations, 2nd, 3rd, 4th, etc. refer to indirect citations. Thus if the same node appears in both first and second place, (e.g. node D) only the first is considered. If the same node appears twice within the same place (node I) it is considered only once.

From a network perspective, citation count refers to a count of the direct ties; this is a major drawback of patent citation analysis. In the present study and in line with other work (Jaffe et al., 2000; Acs et al., 2002; Duguet and MacGarvie, 2005; Peri, 2005;

Criscuolo and Verspagen, 2008; Chang et al., 2009), patent citations are considered a proxy for knowledge flows. Patents are considered an (imperfect) measure of (but a good proxy for) technological innovation (Johnstone et al., 2010), and an incomplete measure of (but a good proxy for) knowledge flows (Criscuolo and Verspagen, 2008). In fact, they capture only those flows that result in novel and patentable technology. Acs et al. (2002) endorse the appropriateness of both patent and innovation counts as reliable measures of knowledge production.

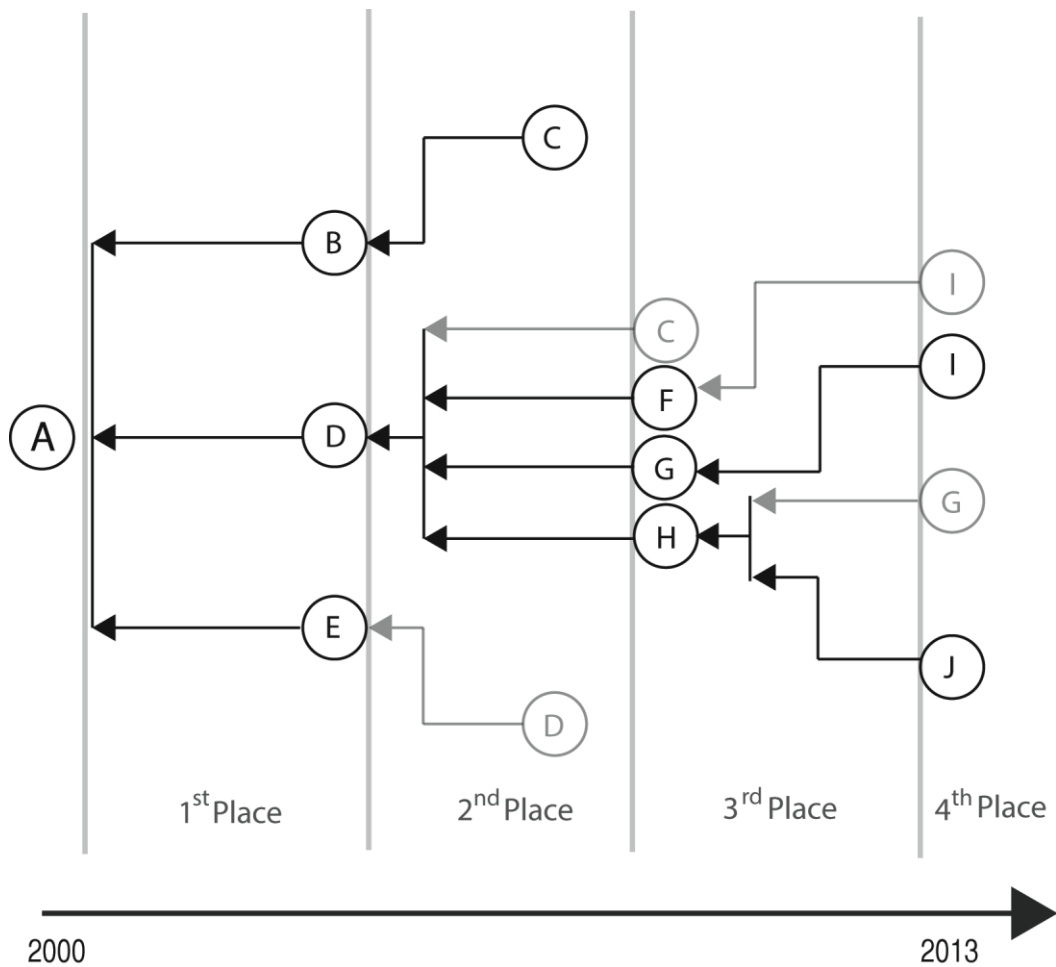


Figure 3. Logical structure of indirect citations

#### Citation distribution

In trying to understand the importance of network citations, it is important to consider their distribution. Some studies highlight some of the structural features of citation networks focusing on the probability distribution function of citations, that is the probability  $P(k^{in})$  that a publication has been cited  $k^{in}$  times. Price (1965) proposes a power law scaling  $P(K^{in}) \sim (K^{in})^{-\gamma}$  with a decaying exponent  $\gamma \approx 3$ , and



Price (1979) theorizes the so-called cumulative advantage mechanism, which refers to the situation in which success breeds success. Redner (1989) confirms power law scaling using a much larger dataset. Others have produced different findings. For instance, Laherrère and Sornette (1989) employ a dataset of the top 1,120 most cited physicists between 1981-1997 and find that the whole distribution of citations is stretched exponential  $P(K^{in}) \sim \exp[-(K^{in})^\beta]$ , with  $\beta \simeq 0.3$ . Subsequently, Redner (2005) analysed all papers published in the 110-year history of the Physical Review and found that the distribution of citations is best fitted by a log-normal distribution.

The lack of consensus on this issue seems to be due to different potential biases, for example, that no account is taken of possible discipline or age-dependence statistics (Radicchi et al., 2012).

### **SNA algorithms: Background**

A popular approach in SNA is citation network analysis to ‘weight’ the importance of individual patents (or journal articles) by counting the number of citations received (Hummon and Doreian, 1989). Patent citations are used to proxy for knowledge flows across a technological field, and for their significance. Patent citation network analysis has been used to trace the development of technological domains and to assess the importance of a patent in a discipline (Batagelj et al., 2006; Fontana et al., 2009; Bekkers and Martinelli, 2012; Epicoco, 2013).

In the case of patent data, patent citation analysis is used to trace ‘technological trajectory’, a concept proposed by Dosi (1982) defined as: ‘A pattern of “normal” problem solving activities on the ground of a technological paradigm’ (Dosi, 1982, p. 152). So far, use of patent citation network analysis to trace the technological trajectory of a field, has been based on the structure of connectivity of the patent citations, using network-based methods and algorithms (Hummon and Doreian, 1989; Batagelj, 2003; Batagelj et al., 2006; Fontana et al., 2009; Bekkers and Martinelli, 2012; Martinelli, 2012). Use of these methods to trace the technological trajectory of a field highlights certain aspects that provide valuable insights into the development and dissemination of the technological innovations characterizing the field. It allows identification of:

- the map of knowledge flows in a sector;
- the milestones related to that field;

- the technological trajectory of a discipline.

The purpose of the present analysis is to trace, visualize and make inferences about the process of technological change in a growing and very interesting industry sector. In the succeeding sections we provide an illustrative example for the proposed algorithm, using the citation network depicted in Figure 2 which includes 10 nodes, Node A, ..., Node J.

## **SNA measure and algorithms for citation networks**

### Centrality measures

Two SNA measures are widely used in analysing citation networks, in-degree and out-degree centrality. The former indicates the number of incoming ties, that is the number of direct citations received by a node in a citations network. The latter refers to the number of outgoing ties and indicates how easily a node can reach other nodes. For example in the citations network depicted in Fig. 2. the degree centrality of Node A is 3.00, and the out-degree is 0.

A third measure, closeness centrality, indicates how easily a node can reach other nodes. A node is considered important if it is relatively close to the all others. Closeness centrality is less intuitive than degree centrality as can be defined as the inverse of the average distance between  $i$  and any other node:  $(n - 1) / \sum_j l(i, j)$ , where  $l(i, j)$  is the number of links in the shortest path between  $i$  and  $j$ . For example the closeness centrality of Node A in Fig. 2 is 0.52 ( $= \frac{9}{17}$ ).

### Hubs and authorities

In the hubs and authority algorithm (Kleinberg, 1999; Brandes and Willhalm, 2002) the authority is the core invention and the hubs are their best development. Hubs and authorities focus on the structure of the citation network and determine its prominent vertices. Hubs and authorities are formal notions of the structural prominence of vertices, identified according to their positions on a graph. This algorithm is developed in the context of journal citations, and the focus on both number of citations and journal prestige represents a step forward.

Hubs and authorities rely on the assumption that, in directed networks, it is possible to identify these two important types of vertices: 'A vertex is a good authority, if it is

pointed to by many good hubs, and it is a good hub, if it points to many good authorities' (Kleinberg, 1999, p. 8).

According to Brandes and Willhalm (2002, p. 3) 'hubs and authorities are eigenvector centralities in the weighted undirected graphs constructed from a directed graph by means of bibliographic coupling and co-citation'. This algorithm was developed in the context of the world wide web; for example, a web page can be considered a citations network. Thus, the two vertices (hubs and authorities) of a page  $p$  can be defined as:  $x(p)$  which is an authority weight and  $y(p)$  its hub weight. The authority weight of page  $p$  is the sum of all hub weights of page  $q$ , for all  $q$  pointing to  $p$ . Then:

$$x(p) = \sum_{q:(q,p)} y(q)$$

The hub weight of page  $p$  is the sum of all authority weights of  $q$  for all  $q$  pointed by  $p$ . Then,

$$y(p) = \sum_{q:(p,q)} x(q)$$

Hubs and authorities focus on the structure of the network and highlight prominent vertices, and seem appropriate for the present research. The knowledge identified shows high levels of basicness, that is, authorities represent core inventions in the network, and hubs their best development.

#### An illustrative example

Consider the network in Figure 2, the most authoritative node is node D with a value equal to 0.90. The highest hub weight is node G with a value equal to 0.57.

#### *Search Path Count*

SPC is used to identify important small sub-networks on the basis of arc weights. SPC reduces the complexity of a large citations network by pinpointing only the 'main flows of knowledge' in which the 'source' is the starting point, that is, a vertex that is cited but cites no other nodes, and the 'sink' is a node that cites other nodes but is not cited. The importance of each citation can be measured by counting the number of times a citation link has been traversed in the path from a set of starting nodes to a set of ending nodes. The main path identified by SPC can be considered

as the most frequent path used to ‘walk’ from the present to the past in a ‘field’ (Calero-Medina and Noyons, 2013).

#### An illustrative example

This algorithm focuses on the connectivity structure more than on nodes, thus we will provide the value of an arc. Consider the arc the arc GD obtains a SPC value of 1, since  $N^-(G)=1$  (paths  $I \rightarrow G$ ); and  $N^+(D) = 1$  (path  $D \rightarrow A$ ).

### **Ordered Weighted Average operator (OWA)**

The family of OWA operators proposed by Yager (1988) includes cumulative operators for membership aggregation. Following this conceptualization, the OWA weighting vector was proposed to introduce the decision maker’s attitude (Yager, 1995), and the OWA operator has been applied in various disciplinary contexts such as decision making under uncertainty (Yager and Kreinovich, 1999), fuzzy information retrieval system (Kacprzyk and Zadrozny, 2001; Herrera-Viedma et al., 2003), e-commerce performance evaluation (Huang et al., 2009) and data mining (Torra, 2004). Emrouznejad and Marra (2014) provided a comprehensive survey and historical development of OWA.

Several approaches have been proposed to obtain associated weights. Filev and Yager (1995) discuss a family of OWA operators which they call exponential OWA operators, and generate weights based on a required degree of ‘orness’. ‘Orness’ refers to the ‘andlike’ or ‘orlike’ aggregation result of an OWA operator (Yager, 1988). This concept has been developed further (Marichal, 1999; Calvo, 2000; Fernández Salido and Murakami, 2003; Yager, 2004).

Along these lines, Wang and Parkan (2005) proposed the minimax disparity method to identify the OWA operator weights using LP under a given level of ‘orness’. In this approach, OWA operator weights are determined by minimizing the maximum difference between two adjacent weights, under a given level of ‘orness’. Wang and Parkan’s model is defined as follows:

#### **Model 1. OWA measure -Wang and Parkan (2005)**

$\min \delta$

$$\text{s. t. } \frac{1}{n-1} \sum_{i=1}^n (n-i)w_i = \alpha$$

$$-\delta \leq w_i - w_{i+1} \leq \delta, \quad i = 1, \dots, n-1,$$

$$\sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n.$$

Amin and Emrouznejad (2006) extended the previous disparity model and proposed the following LP model:

**Model 2. OWA measure - Amin and Emrouznejad (2006)**

$$\min \delta$$

$$\text{s. t. } \frac{1}{n-1} \sum_{i=1}^n (n-i)w_i = \alpha = \text{Orness}(w), \leq \alpha \leq 1,$$

$$-\delta \leq w_i - w_j \leq \delta, i = 1, \dots, n-1, \quad j = i+1, \dots, n,$$

$$\sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n.$$

This model considers the maximum deviation between any adjacent pairs of weights obtained by minimizing the maximum difference between two adjacent weights under a given level of ‘orness’.

A new disparity model that imposes less restriction on the disparity between  $w_j$  and  $w_i$ , was proposed by Emrouznejad and Amin (2010). This model, which can be used to aggregate the preference ranking system, is applied in the present paper and defined as follows:

**Model 3. An OWA measure for preference ranking**

$$\min \sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij}$$

$$\text{s. t. } \frac{1}{n-1} \sum_{i=1}^n (n-i)w_i = \alpha = \text{Orness}(w), 0 \leq \alpha \leq 1$$

$$-\delta_{ij} \leq w_i - w_j \leq \delta_{ij}, \quad i = 1, \dots, n-1 \quad j = i+1, \dots, n,$$

$$\sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n, \delta_{ij} \geq 0, i = 1, \dots, n-1, j = i+1, \dots, n.$$

We use the above model to determine the OWA weight of the patent citation networks studied in this paper. So far, the OWA operator has been applied in business decision making problems, such as the selection of investments (Merigó and Montserrat, 2011) and human resource management (Merigó and Gil-Lafuente, 2011). Citations network is a new domain for the proposed OWA. Here, direct and indirect patent citations are seen as information to aggregate in order to compare networks' nodes.

We used Model 3 to determine the OWA weights associated with direct and indirect citations. As it is shown in (Emrouznejad and Amin, 2010) this model is more appropriate than the standard disparity OWA weights determination model and generates valid weights that accounts for differences among places in a better way than Model 2 (Amin and Emrouznejad 2006). Model (3) complements disparity models, such as Model 1 developed by Wang and Parkan (2005) and Model 2 by Amin and Emrouznejad (2006), rather than superseding them, and specifically it is used for a preference ranking aggregation.

A further justification for using the OWA to rank patent citations is that the decision maker, which in the analysis of patent citation networks might be policy makers or network analysts, wants his or her decision to take account of the role of time as well as indirect citations.

Few existing studies analyse more than three generations of patent citations (von Wartburg et al., 2005; Trajtenberg et al. 2002). In the present study we collect nine generations of patent citations that lead to more complex networks than those studied so far. We refer to these generations as different 'places' (as shown in figure 3). As this complexity increases it becomes necessary to aggregate these places using a method that is able to rank them.

As shown above, one of the challenges issue in studying citation networks is their citation distribution. The proposed OWA allows patents to be ranked according to their indirect citations and considers several places of patents which indicate the evolution of knowledge over time. In this study we used OWA but a more general aggregation function such as Choquet integral-based can also be used, as shown in Fodor et al (1995) any OWA aggregator can be expressed in an equivalent way as a

Choquet integral-based (see also Beliakov and James; 2011, Choquet; 1953-1954, Grabisch and Sugeno; 1992).

An illustrative example

Assume there are 8 patents,  $i = 1, \dots, 8$ . Let us use  $j$  ( $j = 1, \dots, 5$ ) represents the order of citations, i.e.  $j=1$  means number of direct citations,  $j=2$ , means number of indirect citations in the second place, and so on. The numbers of direct and indirect citations for the selected 8 patents are listed in table 1.

**Table 1. Number of patent citations related to selected 8 patents**

| Patents | 1st Place | 2nd Place | 3rd Place | 4th Place | 5th Place |
|---------|-----------|-----------|-----------|-----------|-----------|
| P1      | 4         | 0         | 0         | 0         | 0         |
| P2      | 10        | 62        | 21        | 4         | 1         |
| P3      | 9         | 6         | 1         | 1         | 0         |
| P4      | 8         | 27        | 38        | 30        | 13        |
| P5      | 2         | 6         | 1         | 0         | 0         |
| P6      | 9         | 24        | 35        | 20        | 19        |
| P7      | 3         | 6         | 14        | 56        | 45        |
| P8      | 4         | 3         | 0         | 0         | 0         |

As seen in this table, there are five different places; hence, we determine an OWA vector of five elements, i.e.  $n = 5$ , using the formulation proposed by Emrouznejad and Amin (2010).

**European Patent Office (EPO): Data source**

Our data source is the EPO Worldwide Patent Statistical Database (PATSTAT), which includes patents from 81 national and international patent offices, detailed information on patents published in the EU, and citations from EPO to non-EPO patents, that is, backward and forward citations to other world patents. We identify our patents based on the six categories in the International Patent Classification (IPC) related to the renewable energy sector (wind, solar, geothermal, ocean, biomass, waste), following Johnstone et al. (2010). These categories account for the three generations of technologies that can be distinguished within that sector (International Energy Agency, 2006; Johnstone et al., 2010). The first eight patents analysed in this study deal mainly with inventions related to solar, the most mature of three technologies.

We chose the renewable energy industry because it is a young sector, characterized by a growing patenting activity over the last decade. It is attracting interests from governments since it is seen as playing an important role in national economies (Bergmann et al. 2006). Investments in the renewable energy industry are growing at a fast pace, addressing environmental concerns in developed countries and presenting both challenges and opportunities for emerging countries. They are also having an impact on policy design (Wiser and Pickle, 1998; Inderst et al. 2012). For these reasons we believe that understanding the knowledge dynamics in the renewable industry is very important.

The IPC is a hierarchical classification system applied to published patent documents. We obtained all the patents published in 2000 in the EU, the US and Japan, related to the renewable energy industry. The patenting rate in renewable energy technologies surged in the 2000s (Glachant et al., 2008). Through an iterative process, we identified forward citations received by these patents from other patents published between 2000 and 2013, collecting a total number of 18,135 patents. For the purpose of this study, we present the results for the first eight European patents.

**Table 2. Renewable energy patents published in the EU, US and Japan in 2000**

| <b>Patent Office</b> | <b>Number of patents published</b> |
|----------------------|------------------------------------|
| EU                   | 53                                 |
| US                   | 73                                 |
| Japan                | 437                                |

Data to analyse a citation network has to be relational, that is having ‘Citing’ and ‘Cited’ documents. The originality of the data used in this study consists in having retrieved, through an iterative process, all the citations received (forward citations) by all the patents appearing in each place, as shown in figure 3. Thus, from the original patents published in 2000 in the renewable energy industry we got their forward citations and then the forward citations of the new cohort and so on until 2013. It is worth noting that the number of forward citations follows a scale-free distribution, that is a few patents have most of the links and encounter several places of indirect citations, while the majority of patents have just a few. In doing this, we obtained several indirect citations for each patent.



Identifying the citations received by each patent in our original sample of patents published in 2000 within the specific industry selected, introduces the so called inversion problem. In order to identify all the citations received by a single patent published in 2000, we need to search all the patents granted after 2000, by all patent offices, and in all sectors. Since the EPO database was developed mainly to enable search for backward citations, we have to ‘invert’ the search process. This issue has been discussed by others (see Hall et al., 2001).

## Results and discussions

We ran the analysis for each patent. Hubs and authorities weights and detailed results for all eight patents are available on request but to avoid repetition in this section let us focus on Patent 2 (P<sub>2</sub>) and its network (hereafter NP<sub>2</sub>) only. Results of the OWA scores are described in the last section.

### *Network and connectivity analysis of network built on P<sub>2</sub> (NP<sub>2</sub>)*

P<sub>2</sub> has the label 20001220-00495792. The first part corresponds to date, the second to a publication number. The characteristics of NP<sub>2</sub> are given in Table 3 and a summary of hubs and authorities weights for NP<sub>2</sub> is figured in table 6 and 7. Given the presence of 1 loop we remove it before calculating the network measures.

**Table 3. NP<sub>2</sub> characteristics**

| <b>Number of vertices (n)</b> | <b>99</b>   |
|-------------------------------|-------------|
|                               | <b>Arcs</b> |
| Number of loops               | 1           |
| Number of multiple lines      | 0           |
| Density [loops allowed]       | 0.01        |
| Average Degree                | 2.3         |

*In-degree centrality (Fig. 4, Table 4)* - Fig. 4 depicts NP<sub>2</sub> according to the in-degree centrality measure and the corresponding values are displayed in table 4. According to the in-degree centrality the first patent, the most cited, of NP<sub>2</sub> is patent labelled 20070821-62398637, while P<sub>2</sub> occupies the 4<sup>th</sup> position. P<sub>2</sub> was published in Europe in 2000 by a Japanese applicant Kanepa Corporation with the title “Method of fabricating thin-film photovoltaic module”. The most cited patent has been

published in 2007 in the US with the title “Method of manufacturing thin film photovoltaic modules”, the applicant is the BP Corporation North America Inc.

**Figure 4. In-degree centrality of NP<sub>2</sub>**

**Table 4. Top 10 in-degree and out-degree centrality values of NP<sub>2</sub>**

| Rank | In-degree centrality |       |                   | Title   | Out-degree centrality |       |                   |   |
|------|----------------------|-------|-------------------|---|-----------------------|-------|-------------------|---|
|      | Vertex               | Value | Id<br>(Label)     |   | Vertex                | Value | Id<br>(Label)     | Title   |
| 1    | 3                    | 27    | 20070821-62398637 | Method of manufacturing thin film photovoltaic modules                        | 76                    | 5     | 20120306-75837741 | Methods and related systems for thin film laser scribing devices  |
| 2    | 6                    | 17    | 20080515-29436452 | Method and apparatus for laser beam alignment for solar panel scribing        | 41                    | 4     | 20101216-75416781 | Methods and systems for laser-scribed line alignment  |
| 3    | 4                    | 12    | 20071221-00078925 | Process for laser scribing  | 95                    | 3     | 20121218-74910580 | Process to remove metal contamination on a glass of substrate   |
| 4    | 1 (P2)               | 10    | 20001220-00495792 | Method of fabricating thin-film photovoltaic module                           | 60                    | 2     | 20110621-71003687 | Process and laser scribing  |
| 5    | 2                    | 7     | 20051215-07096762 | Laser structuring for manufacture of thin film silicon solar cells            | 6                     | 2     | 20080515-29436452 | Method and apparatus for laser beam alignment for solar panel scribing  |
| 6    | 60                   | 5     | 20110621-71003687 | Process for laser scribing  | 27                    | 2     | 20100602-73385677 | Laser processing a substrate by a laser processing head along a first processing path, comprises deflecting laser beams on the substrate at a focal point, and determining position of the respective focal points for further processing paths |
| 7    | 13                   | 4     | 20090813-70444497 | Partially transparent solar panel   | 24                    | 2     | 20100311-72844604 | Heating element, and heatable pane comprising a heating element   |
| 8    | 18                   | 4     | 20091231-72194927 | Dynamic scribe alignment for laser scribing, welding or any patterning system | 90                    | 2     | 20120904-72749014 | Laser material removal methods and apparatus  |

|    |   |   |                   |   |    |   |                   |   |
|----|---|---|-------------------|---|----|---|-------------------|---|
| 9  | 8 | 4 | 20090129-69275079 | Verfahren und Vorrichtung zur Laserstrukturierung von Solarzellen             | 44 | 2 | 20101230-75279517 | Device to structure thin-film solar cell module, where structural lines are introduced in module parallel to its transverse edges in rectangular form by laser, includes loading- and unloading station, processing station, and optical device |
| 10 | 7 | 3 | 20080611-19246728 | Electrode configuration for a photovoltaic module with series connected cells | 42 | 2 | 20101216-75427617 | Methods and systems for laser-scribed line alignment  |

**Figure 5. Out-degree centrality values of NP<sub>2</sub>**

Closeness centrality (Fig. 6; Table 5) - P2 is the first among the top 10 patents according to the closeness centrality measure. It means that it is near to the centre of local clusters and is relatively close to all the others. The closeness centrality is calculated considering the numbers of links between each node and all the others in the shortest path. This means that P2 is more close to all the others, so it is easier for it to reach the other nodes than for any others in NP<sub>2</sub>. The concept is more intuitively explained by Fig. 6, which shows P2 (vertex 1) lying at the centre of the surrounding clusters.

**Table 5. Top 10 closeness centrality measures of NP<sub>2</sub>**

| Rank | Vertex | Value | Id (Label)        | Title  |
|------|--------|-------|-------------------|--|
| 1    | 1 (P2) | 0.44  | 20001220-00495792 | Method of fabricating thin-film photovoltaic module                    |
| 2    | 3      | 0.42  | 20070821-62398637 | Method of manufacturing thin film photovoltaic modules                 |
| 3    | 76     | 0.37  | 20120306-75837741 | Method and related systems for thin film laser scribing devices        |
| 4    | 6      | 0.37  | 20080515-29436452 | Method and apparatus for laser beam alignment for solar panel scribing |
| 5    | 60     | 0.37  | 20110621-71003687 | Process for laser scribing   |

|    |    |      |                   |  |
|----|----|------|-------------------|--|
| 6  | 95 | 0.36 | 20121218-74910580 | Process to remove metal contamination on a glass substrate   |
| 7  | 4  | 0.36 | 20071221-00078925 | Process for laser scribing   |
| 8  | 67 | 0.34 | 20110920-71380191 | Method and apparatus for forming separating lines of a photovoltaic module with series-connected cells |
| 9  | 2  | 0.33 | 20051215-07096762 | Laser structuring and manufacture of a thin film silicon solar cells                                   |
| 10 | 7  | 0.31 | 20080611-19246728 | Electrode configuration for a photovoltaic module with series connected cells                          |

**Figure 6. Closeness centrality of NP<sub>2</sub>**

*Authority weights (Fig. 7, Table 6)* - Here, the results of the hubs and authorities algorithm are presented. They identify the first ten most authoritative patents and their best development.

P2 occupies the fifth place in the ranking:

1. The most authoritative patent (20070821-62398637) is the most cited according to the in-degree centrality;
2. The second most authoritative patent (20080515-29436452) was published in 2008, the applicant is a UK company Exitech Ltd, a manufacturer of high-power pulsed laser-based systems for industrial material processing applications. The title of the patent is "Method and apparatus for laser beam alignment for solar panel scribing";
3. The third one (20071221-00078925) belongs to the same owner as the previous patent, it was published in 2007 and deals with a similar technology. The title is "Process for laser scribing".
4. The fourth most authoritative patent (20091231-72194927) was published in 2009 in US, the owner is the company Applied Material Inc. It deals with technology similar to the previous patents, the title is "Dynamic scribing alignment for laser scribing, welding or any patterning system".
5. The fifth one is P2.

**Table 6. The Authoritative patents**

| Rank | Vertex        | Value       | Id (Label)               | Title   |
|------|---------------|-------------|--------------------------|---|
| 1    | 3             | 0.93        | 20070821-62398637        | Method of manufacturing thin film photovoltaic modules                          |
| 2    | 6             | 0.24        | 20080515-29436452        | Method and apparatus for laser beam alignment for solar panel scribing          |
| 3    | 4             | 0.22        | 20071221-00078925        | Process for laser scribing  |
| 4    | 18            | 0.09        | 20091231-72194927        | Dynamic scribing alignment for laser scribing, welding or any patterning system |
| 5    | <b>1 (P2)</b> | <b>0.06</b> | <b>20001220-00495792</b> | Method of fabricating thin-film photovoltaic module                             |
| 6    | 41            | 0.05        | 20101216-75416781        | Methods and systems for laser-scribed line alignment                            |
| 7    | 47            | 0.03        | 20110208-71008390        | Method for producing a photovoltaic module                                      |
| 8    | 2             | 0.01        | 20051215-07096762        | Laser structuring   |

**Figure 7. Authority weights of NP<sub>2</sub>**

*Hub weights (Fig. 8; Table 7)* - In Table 7 we highlight the 10 best developments of the most authoritative patents. Obviously here we will find most recent patents, in fact they have been published mostly in 2011 and 2012. Focusing on the first five hubs, we will find:

1. The first best development (20120306-75837741) is the US patent published as US8129658 (B2) in 2012, owned by Applied Materials Inc with the title “Method and related systems for thin film laser scribing devices”.
2. The second hub (20121218-74910580) has been published in 2012 as US8333843 (B2), owned by the US company Applied Materials Inc, with the title “Process to remove metal contamination on a glass substrate”.
3. The third hub (20110920-71380191) is with the title “Method and apparatus for forming the separating lines of a photovoltaic module with series-connected cells”, published in 2009, owned by German inventor Walter Psyk;
4. The fourth patent (20120904-72749014) is “Laser material removal methods and apparatus”, owned by the company Applied Materials Inc;
5. The fifth hub (20110621-71003687) is US7964820 a patent registered in US by a UK company, the Exitech Ltd, the title is “Process for laser scribing”.

**Table 7. Top 10 Hub patents of NP2**

| Rank | Vertex | Value | Id (Label)        | Title  |
|------|--------|-------|-------------------|--|
| 1    | 76     | 0.29  | 20120306-75837741 | Method and related systems for thin film laser scribing devices  |
| 2    | 95     | 0.26  | 20121218-74910580 | Process to remove metal contamination on a glass substrate   |
| 3    | 67     | 0.21  | 20110920-71380191 | Method and apparatus for forming the separating lines of a photovoltaic module with series-connected cells |
| 4    | 90     | 0.19  | 20120904-72749014 | Laser material removal methods and apparatus   |
| 5    | 60     | 0.18  | 20110621-71003687 | Process for laser scribing   |
| 6    | 66     | 0.18  | 20110920-70963786 | Method for producing a photovoltaic module   |
| 7    | 29     | 0.17  | 20100608-67399182 | Method and system for laser processing targets of different types on a work piece                          |
| 8    | 59     | 0.17  | 20110621-58817246 | Method and apparatus for laser scribing of ultra lightweight semiconductor devices                         |
| 9    | 58     | 0.17  | 20110607-72844891 | Scribe process monitoring methodology  |
| 10   | 12     | 0.17  | 20090610-70400694 | Contacts and module switching from thin layer solar cells to polymer carriers                              |

**Figure 8. Hub weights of NP<sub>2</sub>**

*SPC (Fig. 9; Table 8)* - This section presents the results of the SPC method. It is calculated for paths going from startpoints to endpoints. Figure 9 depicts the 'main path' emerging in NP<sub>2</sub>, it identifies 7 patents. According to the SPC results the technological trajectory goes from P2 to the patent labelled "20120306-75837741" (US8129658 B2), which has been identified previously as the first best hub of the network. The IPC code of this patent suggests that it also belongs to the renewable energy sector (Y02E10/50). Along the trajectory there are five patents already described among the top authoritative patents or in their best development (20070821-62398637, 20110621-71003687, 20101216-75416781, 20091231-72194927, 20120306-75837741).

**Table 8. Vertices on main path SPC [flow] of NP<sub>2</sub>**

| Rank | Vertex              | Cluster | Id (Label)        | Title   |
|------|---------------------|---------|-------------------|---|
| 1    | 1 (P <sub>2</sub> ) | 1       | 20001220-00495792 | Method of fabricating thin-film photovoltaic module                           |
| 2    | 3                   | 1       | 20070821-62398637 | Method of manufacturing thin film photovoltaic modules                        |
| 3    | 18                  | 1       | 20091231-72194927 | Dynamic scribe alignment for laser scribing, welding or any patterning system |
| 4    | 34                  | 1       | 20101028-74934241 | Staggered laser-etch line graphic system, method and articles of manufacture  |
| 5    | 41                  | 1       | 20101216-75416781 | Method and systems for laser-scribed alignment                                |
| 6    | 60                  | 1       | 20110621-71003687 | Process for laser scribing  |
| 7    | 76                  | 1       | 20120306-75837741 | Method and related systems for thin film laser scribing devices               |

**Figure 9. Main path [SPC] of NP<sub>2</sub>**

### An application of the proposed OWA weights in preference ranking

Consider the following example to illustrate the weights generated by the OWA operator. There are eight patents,  $i = 1, \dots, 8$  and  $j = 1, \dots, 9$ , the numbers of direct and indirect citations are listed in Table 9.

**Table 9. Number of citations received by each patent**

|         | 1 <sup>st</sup> | 2 <sup>nd</sup> | 3 <sup>rd</sup> | 4 <sup>th</sup> | 5 <sup>th</sup> | 6 <sup>th</sup> | 7 <sup>th</sup> | 8 <sup>th</sup> | 9 <sup>th</sup> |       |
|---------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-------|
| Patents | Place           | Place           | Place           | Place           | Place           | Place           | Place           | Place           | Place           | Total |
| P1      | 4               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 4     |
| P2      | 10              | 62              | 21              | 4               | 1               | 0               | 0               | 0               | 0               | 98    |
| P3      | 9               | 6               | 1               | 1               | 0               | 0               | 0               | 0               | 0               | 17    |
| P4      | 8               | 27              | 38              | 30              | 13              | 5               | 1               | 0               | 0               | 122   |
| P5      | 2               | 6               | 1               | 0               | 0               | 0               | 0               | 0               | 0               | 9     |
| P6      | 9               | 24              | 35              | 20              | 19              | 15              | 6               | 2               | 1               | 131   |
| P7      | 3               | 6               | 14              | 56              | 45              | 19              | 10              | 5               | 3               | 161   |
| P8      | 4               | 3               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 7     |

Assume  $\alpha = 0.70$  and let's use OWA measure presented in Model 3, hence we have:

|                 | $w_1^*$ | $w_2^*$ | $w_3^*$ | $w_4^*$ | $w_5^*$ | $w_6^*$ | $w_7^*$ | $w_8^*$ | $w_9^*$ |
|-----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| $\alpha = 0.70$ | 0.28    | 0.16    | 0.14    | 0.12    | 0.1     | 0.08    | 0.06    | 0.04    | 0.02    |

And the results ranking the eight patents are given in Table 10.

**Table 10. OWA score for each patents**

| Patents   | P1   | P2   | P3   | P4    | P5   | P6   | P7    | P8  |
|-----------|------|------|------|-------|------|------|-------|-----|
| OWA score | 1.12 | 22.7 | 3.74 | 22.04 | 2.14 | 21.4 | 29.18 | 1.6 |

This means that:  $P7 > P2 > P4 > P6 > P3 > P5 > P1 > P8$ .

The decision to set  $\alpha = 0.70$  is to attribute more importance to citations received in a period close to the publication of the patent, P1....P8. This is consistent with the widely recognized idea that citations distribution follows a power law. In contrast to other works, we consider also the role of the second order of citations received by our original patents. It could be argued that indirect citations that are very far from the first direct citations, no longer reflect the influence of the original patent.

Table 11 summarizes the results for each patent. We only have values for authority weights since our patents P1....P8 are the origins in the network and so cannot be hubs. We also show the closeness centrality values for each patent. The Wilcoxon test results (Table 12-13) confirms the ranking obtain by SNA and OWA are different, as expected.

**Table 11. Measures summary**

| Patents              | P1    | P2     | P3    | P4     | P5     | P6     | P7    | P8     |
|----------------------|-------|--------|-------|--------|--------|--------|-------|--------|
| OWA                  | 1.12  | 22.7   | 3.74  | 22.04  | 2.14   | 21.4   | 29.18 | 1.6    |
| Authority weight     | 1.000 | 0.0627 | 1.000 | 1.000  | 0.0184 | 0.0032 | 0.000 | 0.9997 |
| Hub weight           | 0     | 0      | 0     | 0      | 0      | 0      | 0     | 0      |
| Closeness centrality | 1     | 0.4495 | 1     | 0.3065 | 0.5294 | 0.2729 | 0     | 0.7    |

**Table 12. Wilcoxon signed rank test OWA-authority weights**

| Test statistics            | SNA-OWA |
|----------------------------|---------|
| Asymp. Sig. (2-tailed)     | .012    |
| a. Based on positive ranks |         |

**Table 13. Wilcoxon signed rank test OWA-closeness centrality**

| Test statistics            | SNA-OWA |
|----------------------------|---------|
| Asymp. Sig. (2-tailed)     | .012    |
| a. Based on positive ranks |         |



The technological trajectory of each network is characterized by technology and knowledge coherence since new patents over time develop more complex technologies, that affect industry sectors or lead to new sectors, but are always within the renewable energy industry as demonstrated by the IPC of authority and hub patents.

- NP<sub>1</sub> is a small network of recent patents citing P1, published between 2011 and 2012. The original patent received only 4 direct citations 10 years after its publication. According to the OWA method, it does not receive a significant score, in fact it represent an invention so far not widely exploited with respect to the others. If only its network were analysed, it would represent an authority, but we can see that, compared with the other patents, it does not attract huge attention;
- NP<sub>2</sub> is a bigger network of 99 patents. Its development is characterized by European, US and Japanese company patent ownership. P2, our original patent, is neither the most cited neither nor the best authority in its network. If we were relying only on authority weights, we would focus on the patent published in 2007 (20070821-62398637). We enrich our understanding of NP<sub>2</sub> from the information provided by the OWA scores which provide information on a previous invention. Analysing the technical information provided in each patent we can see that both P2 (20001220-00495792) and the other patent (20070821-62398637) deal with thin-film photovoltaic module;
- NP<sub>3</sub> starts with a EU patent owned by the Japanese Canon KK, whose development includes US and Japanese companies such as the Japanese Sanyo and American Solar Technologies;
- NP<sub>4</sub> is comprised of 123 patents focused mainly on improvements to diode technologies in solar cells. Its development its characterized by domination of US patents and companies – and especially the Emcore Corporation;
- NP<sub>5</sub> is a small network of 10 patents, mainly European, related to with photovoltaics manufacturers;
- NP<sub>6</sub> comprises 132 patents related to solar technologies. One company stands out: the US Vermont Slate & Copper Services;
- NP<sub>7</sub> comprises 163 patents, starting from the one owned by Canon KK, related to solar technologies;
- NP<sub>8</sub> comprises 7 patents published in the US and EU. Development has been influenced mainly by patents belonging to the Japanese company NPC Inc;

P7 does not appear to be an authority patent in the network, nor a very supported one, it is ranked 12th (in-degree = 3.0) according to the in-degree centrality of NP<sub>7</sub>, where the most cited patent achieves a value equal to 48.0, but is a very good vertex along the technological trajectory of its network according to the SPC result. It is the starting point of the main path developed along its network. NP<sub>7</sub> is dominated by one company, the US National Semiconductor Corporation which accounts for three out of 5 hubs. While the hub and authority weights do not attribute a value to P7, it is considered the most important in the OWA ranking. This might be due to the fact that the distribution of citations along the P7 path is different from that of other patents. P7 citations appear mainly after the 3rd place. Its distribution shows that it took longer for its value to be recognised compared to other patents. In line with this argument we can highlight the usefulness of setting orness in the OWA. For example for patents similar to P7, one can suggest an orness level that attribute more importance to citations getting late in time, for example  $\alpha = 0.90$ .

For similar reasons P4 has a higher OWA than P6, despite their receiving a similar number of total citations (direct and indirect). In particular, P6 has a higher number of total citations than P4, but their distribution is different. The indirect citations to P6 came later than those to P4. We chose an 'orness' level that corresponds to attributing more importance to citations received earlier, thus the OWA score for P4 is higher than that for P6.

To summarize, we have shown that the OWA scores enrich our understanding of knowledge evolution within citation networks. The two methods presented, SNA and OWA, may be applied in conjunction to better exploit network dynamics. The main advantage of the OWA operator is that it allows assessment of the role of indirect citations by considering the citations distribution throughout the network and aggregating several generations of patent citations, not so far studied.

## Conclusions

This study analysed patent citations networks in the specific field of renewable energy, based on patents retrieved from the EPO database and published between 2000 and 2013. We applied Hubs and Authorities algorithm to identify the most important contributions.

We have argued that the evolution of knowledge within a citations network should consider both direct and indirect citations. SNA underestimates the role of indirect citations and is not able to provide a network measure for them in citation networks. To address this limitation we integrated this analysis with the application of the OWA weights to show the evolution of some patents over time. Knowledge passes from one node to another within a citations network, so indirect links account for more complex knowledge flows within the network. We proposed the OWA weights to assess the value of indirect citations over time, and propose that nodes with the highest scores survive for longer than those receiving only a high number of direct citations. Our findings reveal that the obtained citation weights differ substantially.

Our interpretation of the OWA weights is in line also with Yager's (1988) original description of the OWA operator as differing from the classical weighted average in the coefficients being associated with ordered position rather than a particular attribute.

Further the proposed model can be applied to measure impact of research using citation analysis. Future work on analysing citations networks using SNA and OWA would contribute to our understanding of the evolution of knowledge, and allow investigation and comparison of the contribution of papers in paper citations networks.

## References

- Acs, Z., Anselin, L., Varga, A., 2002. Patents and innovation counts as measures of regional production of new knowledge. *Research policy* 31(7), 1069–1085.
- Ahuja, G., Morris Lampert, C., 2001. Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal* 22(6-7), 521–543.
- Amin, G.R., Emrouznejad, A., 2006. An extended minimax disparity to determine the OWA operator weights. *Computers & Industrial Engineering* 50(3), 312–316.
- Bacchiocchia, E., Montobbio, F., 2004. EPO vs. USPTO citation lags. 1–19.
- Barberá-Tomás, D., Jiménez-Sáez, F., Castelló-Molina, I., 2011. Mapping the importance of the real world: The validity of connectivity analysis of patent citations networks. *Research Policy* 40(3), 473–486.
- Batagelj, V., 2003. Efficient algorithms for citation network analysis, in: *Institute of Mathematics*. pp. 1–29.
- Batagelj, V., Kejzar, N., Korenjak-Cerne, S., Zavernik, M., 2006. Analyzing the Structure of U.S. Patents Network, in: *Data Science and Classification*. pp. 141–148.

- Bekkers, R., Martinelli, A., 2012. Knowledge positions in high-tech markets: Trajectories, standards, strategies and true innovators. *Technological Forecasting and Social Change* 79(7), 1192–1216.
- Beliakov G., James S., 2011. Citation-based journal ranks: The use of fuzzy measures, *Fuzzy Sets and Systems* 167, 101–1.
- Bergmann, A., Hanley, N., Wright, R., 2006. Valuing the attributes of renewable energy investments. *Energy Policy* 34(9), 1004–1014.
- Borgatti, S.P., Everett, M.G., 2006. A Graph-theoretic perspective on centrality. *Social Networks* 28(4), 466–484.
- Boyack, K., W., Klavans, R. 2010. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404.
- Brandes, U., Willhalm, T., 2002. Visualization of Bibliographic Networks with a Reshaped Landscape Metaphor, in: Ebert, D., Brunet, P., Navazo, I. (Ed), *Joint Eurographics – IEEE TCVG Symposium on Visualization*.
- Calero-Medina, C., Noyons, E. 2008. Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field. *Journal of Infometrics* 2(4), 272–279.
- Calvo, T., 2000. Generation of Weighting Triangles Associated with Aggregation Functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 8(4), 417–451.
- Cantner, U., Graf, H., 2006. The network of innovators in Jena: An application of social network analysis. *Research Policy* 36(4), 463–480.
- Carpenter, M.P., Narin, F., Woolf, P., 1981. Citation rates to technologically important patents. *World Patent Information* 3(4), 160–163.
- Chang, S.B., Lai, K.K., Chang, S.-M., 2009. Exploring technology diffusion and classification of business methods: Using the patent citation network. *Technological Forecasting and Social Change* 76(1), 107–117.
- Choi, J., Yi, S., Lee, K., 2011. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information & Management* 48(8), 371–381.
- Choquet G., 1953–1954. Theory of capacities, *Annales de l'Institut Fourier*, 5, 131–295.
- Criscuolo, P., Verspagen, B., 2008. Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy* 37(10), 1892–1908.
- Dahlin, K.B., Behrens, D.M., 2005. When is an invention really radical? *Research Policy* 34(5), 717–737.
- Dosi, G., 1982. Technological paradigms and technological trajectories. *Research Policy* 11(3), 147–162.
- Duguet, E., MacGarvie, M., 2005. How well do patent citations measure flows of technology? Evidence from French innovation surveys. *Economics of Innovation and New Technology* 14(5), 375–394.
- Emrouznejad, A., Amin, G.R., 2010. Improving minimax disparity model to determine the OWA operator weights. *Information Sciences* 180(8), 1477–1485.

- Emrouznejad, A., Marra, M., 2014. Ordered Weighted Averaging Operators 1988–2014: A citation-based literature survey. *International Journal of Intelligent Systems* (In Press).
- Epicoco, M., 2013. Knowledge patterns and sources of leadership: Mapping the semiconductor miniaturization trajectory. *Research Policy* 42(1), 180–195.
- Fernández Salido, J.M., Murakami, S., 2003. Extending Yager's orness concept for the OWA aggregators to other mean operators. *Fuzzy Sets and Systems* 139(3), 515–542.
- Filev, D., Yager, R.R., 1995. Analytic properties of maximum entropy OWA operators. *Information Sciences* 85(1), 11–27.
- Fontana, R., Nuvolari, A., Verspagen, B., 2009. Mapping technological trajectories as patent citation networks: An application to data communication standards. *Economics of Innovation and New Technologies* 18(4), 311–336.
- Fodor J., Marichal J.-l., Roubens M., 1995. Characterization of the ordered weighted averaging operators, *IEEE Trans. Fuzzy Systems*, 236–240.
- Fujita, K., Kajikawa, Y., Mori, J., Sakata, I., 2014. Detecting research fronts using different types of weighted citation networks. *Journal of Engineering and Technology Management*, 32(April-June), 129–146.
- Garfield, E., 1979. *Citation Indexing*. John Wiley New York.
- Glachant, M., Dechezleprêtre, A., Hascic, I., 2008. Invention and transfer of climate change mitigation technologies on a global scale: A study drawing on patent data. *Fondazione Eni Enrico Mattei Working Papers*, CERNA, Paris.
- Grabisch M., Sugeno M., 1992. Multi-attribute classification using fuzzy integral, *Proceedings of the IEEE International Conference on Fuzzy Systems*, San Diego, 47–54.
- Hall, B., Jaffe, A., Trajtenberg, M., 2001. The NBER patent citation data file: Lessons, insights and methodological tools, in: MIT Press (Ed), *Patents, Citations & Innovations. A Window on the Knowledge Economy*. Cambridge MA, pp. 404–459.
- Hall, B., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. *RAND Journal of economics* 36(1), 16–38.
- Harhoff, D., Scherer, F., Vopel, K., 2003. Citations, family size, opposition and the value of patent rights. *Research Policy* 32(8), 1343–1363.
- Herrera-Viedma, E., Cordon, O., Luque, M., Lopez, A.G., Muñoz, A.M., 2003. A model of fuzzy linguistic IRS based on multi-granular linguistic information. *International Journal of Approximate Reasoning* 34(2-3), 221–239.
- Huang J., X. Jiang, Q. Tang (2009) An e-commerce performance assessment model: Its development and an initial test on e-commerce applications in the retail sector of China. *Information & Management*, 46 (2), 100–108.
- Hummon, N.P., Doreian, P., 1989. Connectivity in a citation network: The development of the DNA theory. *Social Networks* 11(1), 39–63.
- Inderst, G., Kaminker, C., Stewart, F., 2012. Defining and measuring green investments: Implications for institutional investors' asset allocations. *OECD working papers*, 24, 1–55.
- International Energy Agency, 2006. *Renewable energy: R&D priorities. Insights from IEA technology programmes*. Paris.

- Jaffe, A., Trajtenberg, M., Fogarty, M., 2000. The meaning of patent citations: report on the NBER/Case-Western reserve survey of patentees. *American Economic Review* 90(2), 215–218.
- Jaffe, A., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by a patent citations. *The Quarterly Journal of Economics* 108(3), 577–598.
- Johnstone, N., Haščič, I., Popp, D., 2010. Renewable energy policies and technological innovation: Evidence based on patent counts. *Environmental and Resource Economics* 45(1), 133–155.
- Kacprzyk, J., Zadrozny, S., 2001. Computing with words in intelligent database querying: standalone and Internet-based applications. *Information Sciences* 134(1), 71–109.
- Karrer, B., Newman, M., 2009. Random acyclic networks. *Physical Review Letters* 102(128701), 1–4.
- Kessler, M.M. 1963. Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25.
- Kleinberg, J., 1999. Authoritative sources in a hyperlinked environment. *Journal of the Association for Computing Machinery* 46(May), 668–677.
- Laherrère J., Sornette, D., 1998. Stretched exponential distributions in nature and economy: “Fat tails” with characteristic scales. *The European Physical Journal B* 2(4), 525–539.
- Lee, P.-C., Su, H.-N., Wu, F.-S., 2010. Quantitative mapping of patented technology — The case of electrical conducting polymer nanocomposite. *Technological Forecasting and Social Change* 77(3), 466–478.
- Marichal, J., 1999. Aggregation operators for multicriteria decision aid. PhD Thesis, University of Liege, Belgium.
- Martinelli, A., 2012. An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry. *Research Policy* 41(2), 414–429.
- Merigó, J.M., Casanovas, M. 2011. Induced aggregation operators in the Euclidean distance and its application in financial decision making. *Expert Systems with Applications* 38(6), 7603–7608.
- Merigó, J.M., Gil-Lafuente, A. 2011. OWA operators in human resource management. *Economic Computation and Economic Cybernetics Studies and Research* 45(1), 153–168.
- Mina, A., Ramlogan, R., Tampubolon, G., Metcalfe, J.S., 2007. Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research Policy* 36(5), 789–806.
- Newman, M., 2001. The structure of scientific collaboration networks. *Proceedings of the National Academy of Science of the United States of America*.
- Peri, G., 2005. Determinants of knowledge flows and their effect on innovation. *Review of Economics and Statistics* 87(2), 308–322.
- Price de Solla, D. J., 1965. Networks of scientific papers. *Science* 149(3683), 510–515.
- Price de Solla, D. J., 1976. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information* 27(5), 292–306.

- Radicchi, F., Fortunato, S., Vespignani, A., 2012. Citation Networks, in: Scharnhorst, A., Börner, K., Besselaar, P. (Eds), *Models of Science Dynamics*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 233–257.
- Redner, S., 1998. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B* 4(2), 131–134.
- Redner, S., 2005. Citation statistics from 110 years of Physical Review. *Physics Today* 58(6):49–54.
- Rosenkopf, L., Nerkar, A., 2001. Beyond local search: Boundary-spanning, exploration and impact in the optical disk industry. *Strategic Management Journal* 22(4), 287–306.
- Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K., 2009. Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, 60(3), 571–580
- Small, H., 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.
- Torra, V., 2004. OWA Operators in Data Modeling and Reidentification. *IEEE Transactions on Fuzzy Systems* 12(5), 652–660.
- Trajtenberg, M., 1990. A penny for your quotes: Patent citations and the value of innovations. *The RAND Journal of Economics* 21(1), 172–187.
- Trajtenberg, M., Henderson, R., Jaffe, A.B., 2002. University versus corporate patents: a window on the basicness of invention. In: Jaffe A., Trajtenberg, M. (Eds.), *Patents, Citations and Innovations—A Window on the Knowledge Economy*. MIT Press, Cambridge, MA, pp. 51–88, first published in: *Economics of Innovation and New Technology* 5, 19–50.
- Wang, Y.-M., Parkan, C., 2005. A minimax disparity approach for obtaining OWA operator weights. *Information Sciences* 175(1), 20–29.
- Whitley, E.A., Galliers, R.D., 2007. An alternative perspective on citation classics: Evidence from the first 10 years of the European Conference on Information Systems. *Information & Management* 44(5), 441–455.
- Wiser, R., Pickle, S. 1998. Financing investments in renewable energy : the impacts of policy design. *Renewable and Sustainable Energy Reviews* 2(4), 361–386.
- Von Wartburg, I., Teichert, T., Rost, K., 2005. Inventive progress measured by multi-stage patent citation analysis. *Research Policy* 34(10), 1591–1607.
- Yager, R.R., 1988. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics* 18(1), 183–190.
- Yager, R.R., 1995. An approach to ordinal decision making. *International Journal of Approximate Reasoning* 12(3), 237–261.
- Yager, R.R., 2004. Generalized OWA Aggregation Operators. *Fuzzy Optimization and Decision Making* 3(1), 93–107.
- Yager, R.R., Kreinovich, V., 1999. Decision making under interval probabilities. *International Journal of Approximate Reasoning* 22(3), 195–215.

## Appendix: Extracting data from PATSTAT

The dataset used in this study has typical characteristics due to the way data were retrieved from the original source, the EPO PATSTAT. It contains raw data organized in a My-SQL database consisting of 20 tables with rich bibliographic data and citations links among 70 million applications, for more than 80 countries. To extract our data, we query 4 tables linked by the key application identification. This is the application number that identifies univocally each patent.

- 1) The first step consists of retrieving from the database the application id (appln\_id) identifying all patents published in 2000 with one of the IPC codes identifying the renewable energy sector (see Table 13 ). This information is in table 1 and 9 of the database (tls201\_appln; tls209\_appln\_ipc);
- 2) The second step consists of retrieving from table 11 (tls\_211\_pat\_publn) the patent publication identification corresponding to each application id retrieved in the first step;
- 3) The third step consists of retrieving from table 12 (tls12\_citation) patents published after 2000 citing patents published in 2000 within the renewable energy industry. This step is replicated iteratively until zero citations are found.

The final dataset comprises direct and indirect citations to original patents counted and allocated at the corresponding place.

**Table 13. IPC codes for the renewable energy sector**

| Renewable energy technologies | IPC codes – Class | Sub-Class |
|-------------------------------|-------------------|-----------|
| Wind                          | F03D              | 1/00-06   |
|                               | F03D              | 3/00-06   |
|                               | F03D              | 5/00-06   |
|                               | F03D              | 7/00-06   |
|                               | F03D              | 9/00-02   |
|                               | F03D              | 11/00-04  |
|                               | B60L              | 8/00      |
|                               | B63H              | 13/00     |
| Solar                         | F03G              | 6/00-08   |
|                               | F24J              | 2/00-54   |
|                               | F25B              | 27/00B    |
|                               | F26B              | 3/28      |
|                               | H01L              | 31/042    |
|                               | H02N              | 6/00      |
|                               | E04D              | 13/18     |
|                               | B60L              | 8/00      |
| Geothermal                    | F24J              | 3/00-08   |
|                               | F03G              | 4/00-06   |
|                               | H02N              | 10/00     |
| Ocean                         | F03B              | 13/12-24  |
|                               | F03G              | 7/05      |
|                               | F03G              | 7/04      |
|                               | F03B              | 7/00      |



|         |       |         |
|---------|-------|---------|
| Biomass | C10L  | 5/42-44 |
|         | F02B  | 43/08   |
|         | C10L  | 1/14    |
|         | B01J  | 41/16   |
| Waste   | C10L  | 5/46-48 |
|         | F25B  | 27/02   |
|         | F02G  | 5/00-04 |
|         | F23G  | 5/46    |
|         | F012K | 25/14   |
|         | C10J  | 3/86    |
|         | F23G  | 7/10    |
|         | H01M  | 8/06    |